(12)

PARAMETRIC LIKELIHOOD INFERENCE FOR RECORD BREAKING PROBLEMS

by

Bradley P. Carlin

Alan E. Gelfand

TECHNICAL REPORT No. 465

MARCH 9, 1993

Prepared Under Contract

N00014-92-J-1264 (NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

PARAMETRIC LIKELIHOOD INFERENCE FOR RECORD BREAKING PROBLEMS

by

Bradley P. Carlin

Alan E. Gelfand

TECHNICAL REPORT No. 465

MARCH 9, 1993

Prepared Under Contract

N00014-92-J-1264 (NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

Professor Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

Parametric likelihood inference for record breaking problems

BY BRADLEY P. CARLIN

Division of Biostatistics, School of Public Health, University of Minnesota, Box 197 Mayo Memorial Building, Minneapolis, Minnesota 55455-0392, U.S.A.

AND ALAN E. GELFAND

Department of Statistics, University of Connecticut, Box U-120, Storrs, Connecticut 06269-3120, U.S.A.

Summary

In this paper we consider the analysis of record breaking datasets, where only observations that exceed (or only those that fall below) the current extreme value are recorded. Examples of application areas leading to data of this type include industrial stress testing, meteorological analysis, sporting and athletic events, and oil and mining surveys. The inherent missing data structure present in such problems leads to likelihood functions that contain possibly high-dimensional integrals, thus rendering traditional maximum likelihood methods difficult or infeasible. Fortunately, we may obtain arbitrarily accurate approximations to the likelihood function by iteratively applying Monte Carlo integration methods (Geyer and Thompson, 1992). Subiteration using the Gibbs sampler may help to evaluate any multivariate integrals encountered during this process. This approach enables a far more sophisticated set of parametric models than have been applied previously in record breaking contexts. In particular, we illustrate the methodology for a wide array of discrete and continuous distributional settings, and for observations that may be correlated and subject to mean shifts over time. Related issues in model selection and prediction are also addressed. Finally, we present two numerical examples. The first uses a generated dataset exhibiting a high degree of autocorrelation, while the second involves records in Olympic high jump competition.

Some key words: Gibbs sampler; Missing data; Monte Carlo approximant.

1 Introduction

The subject of this paper is estimation and prediction in contexts where data points accumulate sequentially over time, but only those that break the current record (i.e., represent a new maximum or minimum value) are observed. Data of this type arise in a wide variety of practical situations. The history of achievement in sporting and athletic events (such as the times required to run one mile, or top land speeds) is often recorded only in record breaking format. In meteorology, record high temperatures or water levels are sometimes all that are available for a given location. In industrial stress testing, a manufacturer will typically test a series of finished products only up to the current observed minimum strength, rather than simply increasing the pressure on each item until it breaks. Data of this type are closely related to what might be called threshold data, where only observations that exceed (or only those that fall below) a certain level are observed. Examples of this situation include studies of publication bias, drug effectiveness, and patient monitoring in which a physician only sees a patient when the latter is (or believes he is) ill enough to justify the visit.

Datasets of this general class can take many different forms, each requiring its own probability model. Record breaking opportunities may arise in a systematic way (as in an annual auto race) or completely at random (as in a new record for the number of college students in a single phone booth). The former situation seems to call for a discrete time stochastic model; the latter, a continuous time model (although here we might also group the events into convenient time blocks and treat the result as a discrete time series). Another distinction involves whether we have information concerning the failed record breaking attempts or not. In some discrete time settings, we will have such auxiliary information (e.g., for an annual race), but not in most continuous time settings.

Given the form of the dataset in hand, we must adopt realistic assumptions concerning the

nature of the process creating the record breaking attempts. In particular, the assumptions of independence and a constant mean for the sequence giving rise to the records may or may not be appropriate. For example, independence may be reasonable for the sequence leading to record values in destructive stress testing, but not for weekly financial records. When the assumption of independence is not warranted, we may wish to adopt a Markovian dependence structure, but in some cases even this will be too restrictive. Concerning the constant mean hypothesis, records arising from an experienced player's scores in a card game are not likely to shift over time, but we *would* expect such a shift in most athletic competitions due to improvements in equipment, nutrition, conditioning, preparation, and heredity (bearing out the old sports adage, "records were made to be broken").

While there is a large amount of literature on the probability modeling of record breaking data, relatively little exists on the problem of statistical inference in these contexts. Glick (1978) contains an excellent review of the probabilistic work and a brief discussion of tests for randomness in record breaking sequences. Tryfos and Blackmore (1985) discuss the forecasting of future record values given only past records, but only in the case of an independent and identically distributed underlying sequence. Samaniego and Whitaker (1986) focus instead on the problem of inference on the underlying model given the records, but again consider only the independent and identically distributed case, dealing primarily with estimating the mean of a single exponential population. In a second paper, Samaniego and Whitaker (1988) adopt the same framework but with only a nonparametric distributional specification. Smith (1988) retains the independence assumption but drops that of the constant mean, entertaining linear, quadratic, and exponential decay models under normal, Gumbel, and generalized extreme value errors.

In this paper we offer a completely general approach for parametric likelihood inference and prediction in record breaking contexts. The only requirements for our approach to be applicable are

the specification of the joint distribution of the entire data sequence and the index set of the record breaking observations. The general form of the likelihood to be maximized in record breaking problems is laid out in Section 2. Section 3 introduces our Monte Carlo computational approach, and discusses its implementation using Markov chain sampling methods. Section 4 outlines several specific models for record breaking data, in order to give the reader an idea of the generality of the methodology. Section 5 gives two numerical examples illustrating our methodology, the first using an artificial dataset created to exhibit high serial correlation, and the second comprising actual records in Olympic high jump competition.

2 Record breaking likelihoods

Conceptually, record breaking sequences arise within a chronological sequence of events. For a sequence of n events we denote the occurence times by $\{t_1, t_2, \ldots, t_n\}$, and the associated set of measurements by $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$. If the events occur regularly, e.g. daily or annually, we can replace the t's by the index set $\{1, 2, \ldots, n\}$. As noted in Section 1, our formulation presumes that within this sequence we see only the record breaking Y's, but as a result we also know that no records occurred at the unseen Y's. Let $1 = s_1 < s_2 < \cdots < s_r \leq n$ denote the subscripts within $\{t_1, t_2, \ldots, t_n\}$ at which records occurred. Thus we assume a total of r records within the sequence of n events, whence Y_{s_i} denotes the record value associated with s_i. Without loss of generality we assume that larger values break records, i.e., $Y_{s_1} < Y_{s_2} < \cdots < Y_{s_r}$. Hence our dataset is $\{Y_1, s_2, Y_{s_2}, s_3, \ldots, s_r, Y_{s_r}, \text{no records after } t_{s_r}\}$.

We now turn to the likelihood associated with this data. In the existing literature possibly inappropriate simplification with regard to the distribution of the Y_i's has been made, e.g. that they are independent and perhaps identically distributed as well. As suggested in the introduction,

we suspect that in many cases this may not be so, and hence we only assume that $\mathbf{Y}$ has joint distribution $f(\mathbf{y};\theta)$, where θ is a vector of unknown parameters. Therefore the required likelihood can be denoted as $L(\theta; y_1, s_2, \ldots, s_r, y_{s_r})$. Due to the implicit chronology it seems natural to calculate this likelihood as

$$\begin{aligned} & f(y_1;\theta)\, pr(s_2|y_1;\theta) f(y_{s_2}|y_1,s_2;\theta)\cdots pr(s_r|y_1,s_2,\ldots,y_{s_{r-1}};\theta) \\ & \quad \times f(y_{s_r}|y_1,s_2,\ldots,s_r;\theta)\, pr(\text{no records after } t_{s_r}|y_1,s_2,\ldots,y_{s_r};\theta) \end{aligned} \tag{1}$$

To obtain expression (1) we need to compute three types of terms. First consider the term $f(y_{s_j}|y_1,s_2,\ldots,s_j;\theta)$. Define the event $A_i = \{Y_{s_i+1} \le y_{s_i},\ldots,Y_{s_{i+1}-1} \le y_{s_i}\}$ with $A_r = \{Y_{s_r+1} \le y_{s_r},\ldots,Y_n \le y_{s_r}\}$. Then

$$pr(Y_{s_j} \le c|y_1,s_2,\ldots,s_j;\theta) = \frac{pr(y_{s_{j-1}} < Y_{s_j} \le c,\, A_1,A_2,\ldots,A_{j-1}|y_{s_i}, i=1,\ldots,j-1;\theta)}{pr(Y_{s_j} > y_{s_{j-1}},\, A_1,A_2,\ldots,A_{j-1}|y_{s_i}, i=1,\ldots,j-1;\theta)},$$

so that, assuming the derivative exists,

$$f(y_{s_j}|y_1,s_2,\ldots,s_j;\theta) = \frac{f(y_{s_j}|y_{s_i}, i=1,\ldots,j-1;\theta)\, pr(A_1,A_2,\ldots,A_{j-1}|y_{s_i}, i=1,\ldots,j;\theta)}{pr(Y_{s_j} > y_{s_{j-1}},\, A_1,A_2,\ldots,A_{j-1}|y_{s_i}, i=1,\ldots,j-1;\theta)},$$

for $y_{s_j} > y_{s_{j-1}}$. Similarly, since s_j occurs if and only if both of the events $\{Y_{s_j} > y_{s_j}\}$ and A_{j-1} occur, the conditional probability of s_j is

$$pr(s_j|y_1,s_2,\ldots,y_{s_{j-1}};\theta) = \frac{pr(Y_{s_j} > y_{s_{j-1}},\, A_1, A_2,\ldots,A_{j-1}|y_{s_i}, i=1,\ldots,j-1;\theta)}{pr(A_1,A_2,\ldots,A_{j-2}|y_{s_i}, i=1,\ldots,j-1;\theta)}.$$

Lastly,

$$pr(\text{no records after } t_{s_r}|y_1,s_2,\ldots,y_{s_r};\theta) = \frac{pr(A_1,A_2,\ldots,A_r|y_{s_i}, i=1,\ldots,r;\theta)}{pr(A_1,A_2,\ldots,A_{r-1}|y_{s_i}, i=1,\ldots,r;\theta)}.$$

Assembling these pieces we note that a telescoping of terms arises as we calculate the product in (1) so that the likelihood simplifies to

$$f(y_1;\theta)\left\{\prod_{j=2}^{r} f(y_{s_j}|y_{s_i}, i=1,\ldots,j-1;\theta)\right\} pr(A_1,\ldots,A_r|y_{s_i}, i=1,\ldots,r;\theta)\,. \tag{2}$$

A moment's reflection reveals that (2) is, in fact, obvious and might have been written down directly. That is, if we let $\mathbf{U} = (Y_{s_1},\ldots,Y_{s_r})$, $\mathbf{V} = \mathbf{Y}\backslash\mathbf{U}$, and we write $f(\mathbf{y}|\theta) = f(\mathbf{u},\mathbf{v};\theta) = f(\mathbf{u};\theta)f(\mathbf{v}|\mathbf{u};\theta)$, then (2) becomes

$$f(\mathbf{u};\theta)\, pr(\mathbf{V}\in B|\mathbf{u};\theta) = \int_B f(\mathbf{u},\mathbf{v};\theta)d\mathbf{v}\,, \tag{3}$$

where the event $\{\mathbf{V}\in B\} \equiv \{A_1, A_2,\ldots,A_r\}$.

If $\mathbf{Y}$ is a Markov sequence, i.e., $f(\mathbf{y};\theta) = f(y_1;\theta)\prod_{j=2}^{n} f(y_j|y_{j-1};\theta)$ then (2) simplifies to

$$f(y_1;\theta)\left\{\prod_{j=2}^{r} f(y_{s_j}|y_{s_j-1};\theta)\right\}\left\{\prod_{j=1}^{r-1} pr(A_j|y_{s_j}, y_{s_j+1};\theta)\right\} pr(A_r|y_{s_r};\theta)\,. \tag{4}$$

In Section 4 we discuss, in some detail, several specific models for (2) and (4).

3 Maximization of the likelihood

The problem we face can be viewed as one of missing data. Had we observed the v's we would have faced a standard problem, namely maximization of $f(\mathbf{u},\mathbf{v};\theta)$ with respect to θ. Instead, we must maximize a likelihood of the form (2). To describe our approach for obtaining the maximum likelihood estimate of θ, it will be easier to work with the notationally simpler form (3). Given the joint density $f(\mathbf{u},\mathbf{v};\theta)$ and the observations $\mathbf{u} = \mathbf{u}_0$ we need to calculate (3) as a function of θ. Such a function would almost never be available explicitly since it requires an $(n-r)$-dimensional

integration over a constrained region. In general $n - r$ will be large and such integration will defy exact or approximate calculation unless the Y_i are independent, in which case we obtain $n - r$ one-dimensional integrals.

As a result, we are drawn to Monte Carlo approaches for carrying out the integration. In principle, one could attempt a grid search for the maximizing θ, performing a Monte Carlo integration of (3) at each given θ. If the dimension of θ is at all large such searching will be impractical; even for low-dimensional θ the method we now propose will be much faster.

Our objective is to create a Monte Carlo approximant for (3) and subsequently maximize the resulting approximate likelihood. An additional iterative step insures that the likelihood itself is maximized. We obtain our approximant using ideas in Geyer and Thompson (1992) and the associated discussion by Gelfand (1992). Observe that we can write

$$\int_B f(\mathbf{u}_0, \mathbf{v}; \theta) d\mathbf{v} = \left\{ \int_B f(\mathbf{u}_0, \mathbf{v}; \theta_0) d\mathbf{v} \right\} \left\{ \int_B \frac{f(\mathbf{u}_0, \mathbf{v}; \theta)}{f(\mathbf{u}_0, \mathbf{v}; \theta_0)} f(\mathbf{v}|\mathbf{u}_0; \theta_0) d\mathbf{v} \right\} \left\{ \int_B f(\mathbf{v}|\mathbf{u}_0; \theta_0) d\mathbf{v} \right\}^{-1}. \tag{5}$$

Thus if $\mathbf{v}_j^*$, $j = 1, \ldots, M$ are drawn from $g(\mathbf{v}|\mathbf{u}_0; \theta_0)$, the conditional distribution of $\mathbf{V}$ given $\mathbf{u}_0$ and θ_0 restricted to B, a Monte Carlo approximant for (5) is given by

$$\int_B f(\mathbf{u}_0, \mathbf{v}; \theta_0) d\mathbf{v} \times \frac{1}{M} \sum_{j=1}^{M} \frac{f(\mathbf{u}_0, \mathbf{v}_j^*; \theta)}{f(\mathbf{u}_0, \mathbf{v}_j^*; \theta_0)}. \tag{6}$$

Since the integral in this expression is free of θ, an approximate maximum likelihood estimate is obtained by maximizing the summation in (6) with respect to θ. If $f(\mathbf{u}; \theta)$ is available explicitly so that $f(\mathbf{v}|\mathbf{u}; \theta) = f(\mathbf{u}, \mathbf{v}; \theta)/f(\mathbf{u}; \theta)$ is as well, the approximant in (6) can be written equivalently as

$$f(\mathbf{u}_0; \theta) \times \frac{1}{M} \sum_{j=1}^{M} \frac{f(\mathbf{v}_j^*|\mathbf{u}_0; \theta)}{f(\mathbf{v}_j^*|\mathbf{u}_0; \theta_0)}. \tag{7}$$

Expression (7) has computational advantages over (6) under, for example, a Markovian assumption, and is used in the examples in Section 5.

A natural question to ask is how to draw samples from $g(\mathbf{v}|\mathbf{u}_0;\boldsymbol{\theta}_0)$. In special cases, such as multivariate normal models, $f(\mathbf{v}|\mathbf{u}_0;\boldsymbol{\theta})$ will be a standard distribution so that g could be sampled by simple rejection, i.e., retaining $\mathbf{v}_j^*$ drawn from $f(\mathbf{v}|\mathbf{u}_0;\boldsymbol{\theta}_0)$ if and only if it belongs to B. Such sampling will generally be very inefficient, however. An attractive alternative for general f is Markov chain Monte Carlo using the Gibbs sampler (see e.g. Gelfand and Smith, 1990). Implementation requires sampling from the complete conditional distributions arising from $\mathbf{v}$, all of which are proportional to the known joint density $f(\mathbf{u},\mathbf{v};\boldsymbol{\theta}_0)$. In particular if we write $\mathbf{V} = (V_i, \mathbf{V}_{(-i)})$ then we need to sample from $f(v_i|\mathbf{v}_{(-i)}, \mathbf{u}_0;\boldsymbol{\theta}_0)$ restricted to a half interval. If we employ a Metropolis-within-Gibbs algorithm (Müller, 1992) these draws can be made from truncated standard distributions. Such draws may be accomplished using a method suggested in Devroye (1986, p. 38).

Geyer and Thompson (1992) observe that there is gain in iterating the approach. More precisely, starting at some $\boldsymbol{\theta}_0$ if we maximize (6) to obtain $\hat{\boldsymbol{\theta}}$, then we can set $\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}$, redo the maximization resulting in a new $\hat{\boldsymbol{\theta}}$, set $\boldsymbol{\theta}_2$ equal to this new value, and so on. The objective of this iteration is to insure a good Monte Carlo approximant. In practice a few iterations obtain $\boldsymbol{\theta}_l$ in the vicinity of the true $\hat{\boldsymbol{\theta}}$. At this point, one final iteration with M very large will produce an accurate final estimate. A byproduct of this approach is the possibility of approximating the asymptotic covariance of the maximum likelihood estimator. To do so requires calculation (either analytically or numerically) of the Hessian matrix from (6) or (7) at $\hat{\boldsymbol{\theta}}$. We do this in conjunction with our examples in Section 5.

We note that theoretical concerns associated with maximum likelihood estimation, regarding, e.g., existence, uniqueness, consistency, and asymptotic normality, have not been addressed herein. The assumption is that the likelihood under consideration is reasonably well behaved. Remedies for poorly behaved likelihoods are well discussed in the literature and apply here as well.

4 Specific models

4.1 Overview

The Monte Carlo approximant approach of the previous section demonstrates that, under almost any parametric joint density for $\mathbf{Y}$, maximum likelihood estimation given a record breaking sequence can, in principle, be carried out. The goal of this section is to explore more specific models that facilitate calculations and are motivated by the chronological nature of the record breaking process. Rather than attempt any formalization we illustrate with three examples. The first two are fairly general, while the third assumes a Markov Gaussian model.

4.2 Conditionally independent hierarchical models

Suppose that $f(\mathbf{y};\theta)$ arises as $f(\mathbf{y};\theta) = \int f(\mathbf{y}|\mathbf{z};\psi)\, f(\mathbf{z};\eta)d\mathbf{z}$, where $\theta = (\psi,\eta)$. We assume that $f(\mathbf{z};\eta)$ is a proper density over the domain $\mathcal{Z}$ of $\mathbf{z}$ so that $f(\mathbf{y};\theta)$ is proper, and that given $\mathbf{z}$, the Y_i's are independent. Distributional classes of this type have been called conditionally independent hierarchical models (Kass and Steffey, 1989) and offer a rich modeling framework. If we define $h_i(\mathbf{z};\psi) = E(Y_i|\mathbf{z};\psi)$ then $E(Y_i) = E(h_i)$ and $cov(Y_i, Y_j) = cov(h_i, h_j)$. Thus appropriate choice of $f(\mathbf{y}|\mathbf{z};\psi)$ and $f(\mathbf{z};\eta)$ can be made to yield desired model behavior.

If $f(\mathbf{y}|\mathbf{z};\psi)$ and $f(\mathbf{z};\eta)$ form a conjugate pair, marginalization over $\mathbf{z}$ will readily provide $f(\mathbf{y};\theta)$ and we may proceed as in Section 3. If explicit marginalization is not possible, how can we obtain a Monte Carlo approximant to (3)? We can write

$$\begin{aligned}\int_B f(\mathbf{u}_0,\mathbf{v};\theta)d\mathbf{v} \;=\; & \left\{\int_B f(\mathbf{u}_0,\mathbf{v};\theta_0)d\mathbf{v}\right\} \\ & \times\left\{\int_B\int_{\mathcal{Z}} \frac{f(\mathbf{u}_0,\mathbf{v}|\mathbf{z};\psi)f(\mathbf{z};\eta)}{f(\mathbf{u}_0,\mathbf{v}|\mathbf{z};\psi_0)f(\mathbf{z};\eta_0)} f(\mathbf{v}|\mathbf{u}_0,\mathbf{z};\psi_0)f(\mathbf{u}_0|\mathbf{z};\psi_0)f(\mathbf{z};\eta_0)d\mathbf{z}d\mathbf{v}\right\} \\ & \times\left\{\int_B\int_{\mathcal{Z}} f(\mathbf{v}|\mathbf{u}_0,\mathbf{z};\psi_0)f(\mathbf{u}_0|\mathbf{z};\psi_0)f(\mathbf{z};\eta_0)d\mathbf{z}d\mathbf{v}\right\}^{-1}. \qquad (8)\end{aligned}$$

The assumed conditional independence insures that all conditional densities in (8) can be written down immediately. Sup[illegible] $(\mathbf{v}_j^*, \mathbf{z}_j^*)$, $j = 1, \ldots, M\}$ are drawn from the density proportional to $f(\mathbf{v}|\mathbf{u}_0, \mathbf{z}; \psi_0) f(\mathbf{u}_0|\mathbf{z}; \psi_0) f(\mathbf{z}; \eta_0)$ restricted to $B \times \mathcal{Z}$. Then an approximant to (8) is given by

$$\int_B f(\mathbf{u}_0, \mathbf{v}; \theta_0) d\mathbf{v} \times \frac{1}{M} \sum_{j=1}^{M} \frac{f(\mathbf{u}_0, \mathbf{v}_j^*|\mathbf{z}_j^*; \psi) f(\mathbf{z}_j^*; \eta)}{f(\mathbf{u}_0, \mathbf{v}_j^*|\mathbf{z}_j^*; \psi_0) f(\mathbf{z}_j^*; \eta_0)}, \tag{9}$$

which is a minor extension of equation (6). The required sampling of $\mathbf{v}$ and $\mathbf{z}$ over $B \times \mathcal{Z}$ may be carried out by extending the Gibbs sampler as follows. Given $\mathbf{z}$ draw $\mathbf{V}$ from $f(\mathbf{v}|\mathbf{u}_0, \mathbf{z}; \psi_0)$ restricted to B using the complete conditional distributions for V_i, which are free of $\mathbf{u}_0$ and $\mathbf{v}_{(-i)}$ under conditional independence. Given $\mathbf{v}$ draw $\mathbf{Z}$ using the complete conditionals for the Z_i, which are proportional to the nonnormalized form $f(\mathbf{v}|\mathbf{u}_0, \mathbf{z}; \psi_0) f(\mathbf{u}_0|\mathbf{z}; \psi_0) f(\mathbf{z}; \eta_0)$. The Z_i are treated as missing data, just like the V_i.

4.3 Moving window sum processes

A natural extension of the case of independent Y_i's is to envision them arising as observations from a moving window sum of independent variables. We illustrate for a window of size two. Suppose that $Z_0, Z_1, \ldots, Z_n$ are independent with $Z_i \sim f_i(\,\cdot\,; \theta)$. Let $Y_i = Z_i + Z_{i-1}$. Then the joint distribution of $\mathbf{Y}$ can be straightforwardly written down. Moreover $E(Y_i) = E(Z_i) + E(Z_{i-1})$ and $cov(Y_{i-1}, Y_i) = var(Z_{i-1})$, so that appropriate choices of the f_i will provide desired model behavior.

As a concrete example, suppose a new species is introduced into an area and thereafter seasonal population counts are of interest but, in fact, over the course of say n seasons only the record breaking seasons and counts are recorded. Typically such counts are modeled as Poisson variates, but here it might be inappropriate to assume they are independent. Suppose we let Z_i be inde-

pendent $Poisson(\lambda_i)$ random variables for $i = 0, 1, \ldots, n$ and set $Y_i = Z_i + Z_{i-1}$. Then clearly $Y_i \sim Poisson(\lambda_i + \lambda_{i-1})$, $cov(Y_i, Y_{i-1}) = \lambda_{i-1}$, and $E(Y_i|Y_{i-1}) = \lambda_i + \lambda_{i-1}Y_{i-1}/(\lambda_{i-1} + \lambda_{i-2})$, linear in Y_{i-1}. To create a drift in $E(Y_i)$ we could take the λ_i to be specified parametric functions. If λ_i is linear in i, we have $E(Y_i)$ linear in i; if λ_i is exponential in i, then $\log E(Y_i)$ is linear in i.

4.4 Markov Gaussian models

A Markovian assumption for a record breaking sequence seems a plausible yet relatively simple extension from independence. Recall that under such an assumption the likelihood simplifies a bit as in expression (4). In particular suppose that all marginal and conditional densities from $f(\mathbf{y}; \theta)$ can be obtained explicitly, as in the case of $f(\mathbf{y}; \theta)$ multivariate normal. Let $\mathbf{W}_i = (Y_{s_i+1}, \ldots, Y_{s_{i+1}-1})$, $i = 1, \ldots, r-1$, and $\mathbf{W}_r = (Y_{s_r+1}, \ldots, Y_n)$. Then the $\mathbf{W}_i$ are conditionally independent given u yielding (4). Since $pr(\mathbf{V} \in B|\mathbf{u}_0, \theta)$ can be written as a product of r terms, an approximant for each term can be created using low-dimensional Gibbs samplers or some other numerical integration technique, rather than requiring one fully $(n-r)$-dimensional sampler. Expression (7) would be recast as a product of sums utilizing the conditional distributions $f(\mathbf{w}_i|\mathbf{u}_0; \theta)$, $i = 1, \ldots, r$.

Suppose, in fact, that events occur at regular intervals and the process is first order Gaussian, i.e. $Y_i - \mu_i = \rho(Y_{i-1} - \mu_{i-1}) + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$ and $\mu_i = E(Y_i)$ is a specified parametric function. As in the previous example the μ_i can reflect drift; for instance, $\mu_i = \alpha + \beta i$ is used in the examples of Section 5. Clearly $var(Y_i) = \sigma^2/(1-\rho^2)$ and $cov(Y_i, Y_{i+j}) = \sigma^2\rho^j/(1-\rho^2)$. Such a model extends Smith (1988) and yields routine distribution theory. Alternatively, Markov Gaussian models may be created through the inverse covariance matrix. Whittaker (1990, Chapters 5 and 6) presents a very readable discussion of this approach. In particular he shows that if only the diagonal and first off-diagonal terms in the inverse of the covariance matrix are nonzero, then the

joint distribution is a Markov Gaussian model.

A generalization in the spirit of Subsection 4.2 assumes that the μ_i are random. The idea is that randomly arising larger μ_i will encourage the breaking of records. If so then we might write the model in hierarchical form as $f(y|\mu;\rho,\sigma^2)f(\mu;\eta)$. Note however that we do not have a conditionally independent hierarchical model. An illustration would be the first order dynamic, or state-space, model (see e.g. Carlin, Polson and Stoffer, 1992). If we can explicitly marginalize over μ we will find ourselves with a Markov Gaussian model again. If not, we can create an approximant similar to (9).

A final related point here concerns the extension of such models to a continuous time process $\{Y(t) : t \geq 0\}$, where events occur at times $t_1, \ldots, t_n$ resulting in $Y(t_1), \ldots, Y(t_n)$. Here a standard theorem in stochastic processes (see e.g. Breiman, 1986, p. 289) notes that a Gaussian stationary process is Markov if and only if its autocovariance function $\Gamma(t)$ is of the form $\delta\gamma^{|t|}$, $0 < \gamma < 1$. In other words, the only stationary Markov Gaussian process is of the form we have just described.

5 Numerical examples

5.1 Simulated high correlation data

To illustrate the performance of the methodology for record breaking data exhibiting high autocorrelation, consider the $n = 50$ simulated y_i values given in Table 1. These data were generated according to the linear first order Gaussian model introduced in Subsection 4.4, where we let $\alpha = 0$, $\beta = 1$, $\sigma = 1.4$, $\rho = 0.8$, and set $y_1 = 1$. The y_i values are displayed graphically as dots in Figure 1(a); the $r = 34$ record breaking observations are boxed for easier identification. Fitting a simple linear regression model to the full 50-point dataset, we see in Figure 1(b) the wave pattern often present in the residual plot for serially correlated data, and a clear positive trend in the plot of each residual versus the immediately preceeding residual in Figure 1(c). Both of these diagnostics

suggest high positive correlation for our generated dataset.

i	no record	y_i	i	no record	y_i	i	no record	y_i	i	no record	y_i
1		1.00	13		14.81	26	*	27.71	39		41.69
2		1.02	14		19.14	27	*	28.44	40		41.81
3		1.18	15		21.59	28	*	28.41	41		41.85
4		3.07	16		23.63	29		30.06	42		43.63
5		5.32	17		24.49	30		31.08	43		43.73
6	*	3.67	18		25.01	31		33.68	44		47.70
7	*	4.22	19	*	24.85	32		34.58	45		47.92
8		8.55	20	*	23.12	33	*	33.08	46		49.37
9	*	8.50	21	*	24.95	34	*	30.19	47		52.01
10		9.37	22	*	24.83	35	*	31.40	48		53.25
11		10.95	23		25.93	36	*	33.00	49		56.84
12		12.33	24		28.85	37		35.72	50	*	54.92
			25	*	28.50	38		35.73			

Table 1: Simulated data having $\rho = 0.8$

The magnitude of the Pearson correlation present in the lagged residual plot, 0.83, indicates the severity of the autocorrelation, but its value is a bit misleading since this statistic does not actually estimate the parameter ρ in an AR(1) model with a time trend. Instead, we might look at the differenced series $D_t = Y_t - Y_{t-1}$, $t = 2, \ldots, 50$, and observe that $var(D_t) = 2\sigma^2/(1+\rho)$ and $cov(D_t, D_{t-1}) = -[\sigma^2(1-\rho)]/(1+\rho)$, so that $corr(D_t, D_{t-1}) = -(1-\rho)/2$. Hence if $\hat{C}$ is the sample estimate of this correlation, we obtain the crude point estimator $\hat{\rho} = 1 + \min(0, 2\hat{C})$. Figure 1(d) plots the (D_t, D_{t-1}) pairs for our dataset; the resulting $\hat{\rho}$ equals 0.901. Thus we have substantial evidence that a model incorporating ρ will substantially improve our inference.

Of course, our models will not analyze the full data as above, but only the 68% of the data that constitute record breaking observations. Notice that there are three gaps of length four (i = 19–22, 25–28, and 33–36), one gap of length two (i = 6–7), and two gaps of length one (i = 9 and 50). We will use our Monte Carlo algorithm to find the maximum likelihood estimate of $\theta = (\alpha, \beta, \sigma, \rho)$ given the observed data $\mathbf{u}_0 = (y_{s_1}, \ldots, y_{s_{34}})$. We will compare the fit of this model to that of a

reduced model where we ignore autocorrelation by insisting that $\rho = 0$.

Using notation suggested in Subsection 4.4, we write $\mathbf{w}_1 = (y_6, y_7)'$, $\mathbf{w}_2 = (y_{19}, y_{20}, y_{21}, y_{22})'$, $\mathbf{w}_3 = (y_{25}, y_{26}, y_{27}, y_{28})'$, and $\mathbf{w}_4 = (y_{33}, y_{34}, y_{35}, y_{36})'$, so that $\mathbf{v} = (y_9, y_{50}, \mathbf{w}_1', \mathbf{w}_2', \mathbf{w}_3', \mathbf{w}_4')'$. Then the likelihood (4) for our dataset takes the form

$$\begin{aligned} L(\theta; \mathbf{u}_0) = f(y_1; \theta) \left\{ \prod_{j=2}^{7} f(y_{s_j} | y_{s_{j-1}}; \theta) \right\} \left\{ \int_{-\infty}^{y_8} f(y_9 | y_8, y_{10}; \theta) dy_8 \right\} \left\{ \int_{-\infty}^{y_{49}} f(y_{50} | y_{49}; \theta) dy_{50} \right\} \\ \times \left\{ \int_{-\infty}^{y_5} \int_{-\infty}^{y_8} f(\mathbf{w}_1 | y_5, y_8; \theta) d\mathbf{w}_1 \right\} \left\{ \int_{-\infty}^{y_{18}} \int_{-\infty}^{y_{18}} \int_{-\infty}^{y_{18}} \int_{-\infty}^{y_{18}} f(\mathbf{w}_2 | y_{18}, y_{23}; \theta) d\mathbf{w}_2 \right\} \\ \times \left\{ \int_{-\infty}^{y_{24}} \int_{-\infty}^{y_{24}} \int_{-\infty}^{y_{24}} \int_{-\infty}^{y_{24}} f(\mathbf{w}_3 | y_{24}, y_{29}; \theta) d\mathbf{w}_3 \right\} \left\{ \int_{-\infty}^{y_{32}} \int_{-\infty}^{y_{32}} \int_{-\infty}^{y_{32}} \int_{-\infty}^{y_{32}} f(\mathbf{w}_4 | y_{32}, y_{37}; \theta) d\mathbf{w}_4 \right\}, \end{aligned}$$

Since we have assumed a first order linear Gaussian model, the distribution for each observed record given the one immediately preceding, $f(y_{s_i} | y_{s_{i-1}}; \theta)$, is readily available from standard multivariate normal theory. Similarly, the required conditional distributions for the gaps $y_9, y_{50}, \mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ and $\mathbf{w}_4$ are also available as normals, completing the likelihood specification. Thus a Monte Carlo approximant of the form in (7) with the Markovian simplifications discussed in Subsection 4.4 is convenient.

To carry out the required $\mathbf{v}_j^*$ sampling, we first note that $y_{9,j}^*$ and $y_{50,j}^*$ values may be generated directly from their (suitably truncated) complete conditional distributions, obtained from standard multivariate normal theory as

$$\begin{aligned} f(y_9 | y_8, y_{10}; \theta) &\propto N(\mu_9, \sigma^2/(1+\rho^2)) I_{(-\infty, y_8)}(y_9), \quad \text{and} \\ f(y_{50} | y_{49}; \theta) &\propto N(\mu_{50}, \sigma^2) I_{(-\infty, y_{49})}(y_{50}), \end{aligned} \tag{10}$$

where $\mu_9 = \alpha + 9\beta + \frac{2\rho}{1+\rho^2}\left(\frac{y_8+y_{10}}{2} - \alpha - 9\beta\right)$ and $\mu_{50} = \alpha + 50\beta + \rho(y_{49} - \alpha - 49\beta)$. For the missing data lying in gaps of length greater than 1, however, we resort to Markov chain Monte Carlo methods to obtain the necessary samples. For each such missing y_i, the complete conditional

density to be used for generation of the y_{ij}^*'s is of the form

$$f(y_i|y_{i-1}, y_{i+1}; \theta) \propto N(\mu_i, \sigma^2/(1+\rho^2))I_{(-\infty, y_{s_{i-1}})}(y_i), \tag{11}$$

where $\mu_i = \alpha + \beta i + \frac{2\rho}{1+\rho^2}\left(\frac{y_{i-1}+y_{i+1}}{2} - \alpha - \beta i\right)$, $y_{s_{i-1}}$ is the most recent record value, and it is understood that either or both of the conditioning values y_{i-1} and y_{i+1} may themselves be Monte Carlo samples if they too correspond to non-record values. In our implementation, we ran M parallel sampling chains for N "burn-in" iterations to reach the chain's ergodic distribution, retaining only the N^{th} value from each chain. While somewhat wasteful, this approach was an easy way to obtain independent iterates in a situation where the required generation was inexpensive. We took $N = 20$, a conservative burn-in value based on our experience with normal sampling models. In less regular modeling scenarios, a monitoring diagnostic may help select the proper value of N. Under the parallel sampling approach, the recent papers by Gelman and Rubin (1992), Ritter and Tanner (1992), and Roberts (1992) are particularly helpful in this regard.

Given the sampled values $\{\mathbf{v}_j^* = (y_{0j}^*, y_{50,j}^*, \mathbf{w}_{1j}^*, \mathbf{w}_{2j}^*, \mathbf{w}_{3j}^*, \mathbf{w}_{4j}^*),\ j = 1, \ldots, M\}$, the Monte Carlo approximant (7) is easily computed. As mentioned in Section 3, our algorithm uses a small number of iterations to update this θ_0 value before the final maximization. Our program was written in FORTRAN and called the IMSL routine DBCONF, a quasi-Newton algorithm employing a finite difference gradient, to perform the necessary maximizations.

Using $M = 10,000$ replications on the third and final iteration of the algorithm we obtained the full model maximum likelihood estimate $\hat{\theta} = (-0.008, 1.073, 1.706, 0.786)$. A numerically computed Hessian produced the asymptotic standard deviation vector (2.001, 0.068, 0.220, 0.091); the estimated correlations between the elements of $\hat{\theta}$ were all negligible with the exception of that between $\hat{\alpha}$ and $\hat{\beta}$, –0.85. Fitting the simple trended AR(1) model to the entire set of $n = 50$ observations

results in the point estimate (0.208, 1.068, 1.632, 0.819) and associated standard deviation vector (2.165, 0.072, 0.168, 0.078). The estimation that uses record data only appears to be degraded very little, despite the loss of nearly one third of the observations.

Repeating the algorithm using the same number of iterations and replications for the reduced model ($\rho = 0$) gave $\hat{\alpha} = -0\ 106$, $\hat{\beta} = 1.055$, and $\hat{\sigma} = 3.043$. Figure 1(a) shows that the trend lines obtained from the full and reduced models are virtually indistinguishable. Still, the χ^2 likelihood ratio statistic equals 36.76 on 1 degree of freedom, and this in turns leads to an Akaike information criterion of 35.76 and a Bayesian information (Schwarz) criterion of 33.23 – the latter suggesting a Bayes factor in favor of the full model of over 16 million!

Clearly all of the above model choice criteria confirm that the full model is vastly superior, but apparently its value lies primarily in its increased precision, indicated by its much smaller $\hat{\sigma}$ value. Since this added precision should translate into better predictive ability, we decided to investigate further using a bootstrap approach. We drew $y^*_{50,j}, \ldots, y^*_{60,j}$ from the fitted full and reduced models for $j = 1, \ldots, 2000$, being careful to constrain $y^*_{50,j}$ to be less than y_{49}, as was observed in the original dataset. Figure 2(a) then plots the 5^{th}, 50^{th} and 95^{th} percentiles of the bootstrap distribution over time. We see that the full model is indeed more precise, though its advantage gradually diminishes. The full model expects slightly larger future observations on the average due to the recent history of records at y_{37} through y_{49}. The compression of the 95^{th} percentile for $y^*_{50,j}$ is apparently due to the restriction that it not exceed y_{49} – an extra bit of information not usually available in a truly predictive setting.

Finally, Figure 2(b) gives the histogram of the bootstrap distribution of the waiting time until the first record after y_{49} under the full model. Counting the known non-record value at $t = 50$, the model suggests substantial probabilities of gaps of length 4 or 5, and a nonnegligible chance of a gap as long as 10 time points. Again, this behavior is understandable in light of the high estimated

autocorrelation combined with the long string of records recently observed to have ended; indeed, such gaps have occurred in our sample.

5.2 Olympic high jump data

As a second illustration, consider the data displayed in Table 2. These are the record breaking Olympic high jumps since 1896, as presented in the *World Almanac and Book of Facts* (1989). Besides being a prototype for many sports history datasets of this type, this dataset is interesting because it contains two distinct types of missing data. First, no record breaking high jump occurred at the Olympics in the years 1904, 1920, 1928, 1932, 1948, 1972, and 1984. Second, no record occurred in the years 1916, 1940, and 1944 because the Olympics themselves were cancelled due to the intervening world wars. Our likelihood must reflect this distinction between the failures and the cancellations.

j	s_j	year	record (in.)	athlete (country)
1	1	1896	71.25	Ellery Clark (US)
2	2	1900	74.80	Irwing Baxter (US)
3	4	1908	75.00	Harry Porter (US)
4	5	1912	76.00	Alma Richards (US)
5	8	1924	78.00	Harold Osborn (US)
6	11	1936	80.00	Cornelius Johnson (US)
7	15	1952	80.32	Walter Davis (US)
8	16	1956	83.50	Charles Dumas (US)
9	17	1960	85.00	Robert Shavlakadze (USSR)
10	18	1964	85.75	Valery Brumel (USSR)
11	19	1968	88.25	Dick Fosbury (US)
12	21	1976	88.50	Jacek Wszoia (Poland)
13	22	1980	92.75	Gerd Wessig (E. Germany)
14	24	1988	93.50	Guennadi Avdeenko (USSR)

Table 2: Olympic High Jump Records, 1896–1988

Since world-class athletes typically compete in more than one year's Olympics, and perhaps since the athletes of a few countries seem to dominate this event, we might wish to fit a model

for dependent data. With $r = 14$ broken records in only 21 Olympic attempts, there is also a clear need to model an increasing mean over time. Because this is again a discrete time dataset, we shall attempt to fit the normal linear AR(1) model presented in Subsection 4.4. For simplicity we let $\mu_i = \alpha + \beta i = \alpha + \beta(year - 1892)/4$, and investigate the distribution of the waiting time until the next record. Basak and Bagchi (1990) used Laplace's method to estimate the predictive distribution of the magnitude of the next record given the 14 records in this dataset, but fit only a simple model that assumed uncorrelated observations having a constant mean over time. Their analysis also ignored the failures and cancellations, and thus discarded the information carried by the failed attempts. Under this model the cancellations *may* be ignored, since they cannot be connected to any observed data and thus do not affect the likelihood.

Writing $\mathbf{w} = (y_9, y_{10})'$ so that $\mathbf{v} = (y_3, y_{20}, y_{23}, y_7, y_{14}, \mathbf{w}')'$, the likelihood is given by

$$L(\theta; \mathbf{u}_0) = f(y_1; \theta) \left\{ \prod_{j=2}^{r} f(y_{e_j} | y_{e_j - 1}; \theta) \right\} \left\{ \prod_{j \in F_1} \int_{-\infty}^{y_{j-1}} f(y_j | y_{j-1}, y_{j+1}; \theta) dy_j \right\}$$
$$\times \left\{ \int_{-\infty}^{y_8} f(y_7 | y_5, y_8; \theta) dy_7 \right\} \left\{ \int_{-\infty}^{y_{11}} f(y_{14} | y_{11}, y_{15}; \theta) dy_{14} \right\} \left\{ \int_{-\infty}^{y_8} \int_{-\infty}^{y_8} f(\mathbf{w} | y_8, y_{11}; \theta) d\mathbf{w} \right\},$$

where $F_1 = \{3, 20, 23\}$. All of the distributions in this expression are available as univariate normals, except for the bivariate normal distribution of $\mathbf{w}$.

Turning to our sampling-based implementation, we again work with expression (7), which requires simulated y_{ij}^* values for the seven observed failures. Except for the back-to-back failures in 1928 and 1932, all of these represent gaps of length 1 and thus may be generated without the use of Gibbs sampling from complete conditional distributions similar to those displayed in equation (10). Of course when one of these failures abuts a cancellation (as in 1920 and 1948), there are slight modifications to the mean and variance to reflect the fact that the adjacent record is more than one position away, but the associated conditional normal calculations are still routine. For the single gap of length two, we used the Gibbs sampler with $N = 20$ to obtain $y_{9,j}^*$ and $y_{10,j}^*$ values generated

from complete conditionals of the form given in equation (11). No modifications are needed in this case as both y_8 and y_{11} are observed records.

Once in possession of the sampled values $\{\mathbf{v}_j^* = (y_{3j}^*, y_{20,j}^*, y_{23,j}^*, y_{7j}^*, y_{14,j}^*, \mathbf{w}_j^*),\ j = 1, \ldots, M\}$ we may evaluate the Monte Carlo approximant to the likelihood (7), which is again routine using multivariate normal theory. Generating $M = 10,000$ replications at the final iteration, this dataset required only $N = 3$ iterations to produce the maximum likelihood estimate $\hat{\theta} = (70.037, 0.894, 1.697, 0.334)$, with associated asymptotic standard deviation vector (1.384, 0.084, 0.368, 0.372). Appealing to the asymptotic normality of the maximum likelihood estimator, the data strongly suggest an increase of nearly 1 inch in the best Olympic high jump every four years, but offer only mild evidence of a positive correlation amongst these quadrennial performances. Repeating the calculations for the reduced model having $\rho = 0$, we obtained $\hat{\alpha} = 70.072$, $\hat{\beta} = 0.881$, and $\hat{\sigma} = 1.869$, estimates which show little movement from those in the full model. The single degree of freedom χ^2 likelihood ratio statistic between these two models is 1.91, implying a p-value of 0.168. The Akaike criterion is 0.91 (a slight preference for the full model) while the Schwarz criterion is –0.73, for an approximate Bayes factor of 0.694 = 1/1.44 (a slight preference for the reduced model). Thus the data are inconclusive on the issue of whether to include ρ in the model or not.

Figure 3(a) addresses the prediction question, again through a bootstrapping approach. While the fitted full model has been used in this analysis, the near linearity of all three lines in this plot indicates only a small gain in predictive precision over what might be expected from the simple uncorrelated model. Finally, Figure 3(b) shows a nearly linear decline in the probabilities of the waiting time distribution, with a new record almost certain to have occurred within the next five Olympic meets. The 37% predicted chance of a record-breaking high jump at the 1992 Olympics seems consistent with the overall observed proportion of records in back-to-back competitions.

References

BASAK, P. & BAGCHI, P. (1990). Application of Laplace approximation to record values. *Communications in Statistics, Part A - Theory and Methods* **19**, 1875–1888.

BREIMAN, L. (1986). *Probability and Stochastic Processes: With a View Toward Applications.* Palo Alto, California: Scientific Press.

CARLIN, B.P., POLSON, N.G. & STOFFER, D.S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *J. Am. Statist. Assoc.* **87**, 493–500.

DEVROYE, L. (1986). *Non-Uniform Random Variate Generation.* New York: Springer-Verlag.

GELFAND, A.E. (1992). Discussion to "Constrained Monte Carlo maximum likelihood for dependent data," by C.J. Geyer and E.A. Thompson. To appear *J. R. Statist. Soc.* B.

GELFAND, A.E. & SMITH, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* **85**, 398–409.

GELMAN, A. & RUBIN, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). To appear *Statistical Science.*

GEYER, C.J. & THOMPSON, E.A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with discussion). To appear *J. R. Statist. Soc.* B.

GLICK, N. (1978). Breaking records and breaking boards. *American Mathematical Monthly* **85**, 2–26.

KASS, R.E. & STEFFEY, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Am. Statist. Assoc.* **84**, 717–726.

MÜLLER, P. (1992). A generic approach to posterior integration and Gibbs sampling. To appear *J. Am. Statist. Assoc.*

RITTER, C. & TANNER, M.A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the griddy Gibbs sampler. To appear *J. Am. Statist. Assoc.*

ROBERTS, G.O. (1992). Convergence diagnostics for the Gibbs sampler. To appear in *Bayesian Statistics 4*, Ed. J.M. Bernardo, J.O. Berger, D.V. Lindley and A.F.M. Smith. Oxford: University Press.

SAMANIEGO, F.J. & WHITAKER, L.R. (1986). On estimating population characteristics from record-breaking observations. I. Parametric results. *Naval Research Logistics Quarterly* **33**, 531–543.

SAMANIEGO, F.J. & WHITAKER, L.R. (1988). On estimating population characteristics from record-breaking observations. II. Nonparametric results. *Naval Research Logistics Quarterly* **35**, 221–236.

SMITH, R.L. (1988). Forecasting records by maximum likelihood. *J. Am. Statist. Assoc.* **83**, 331–338.

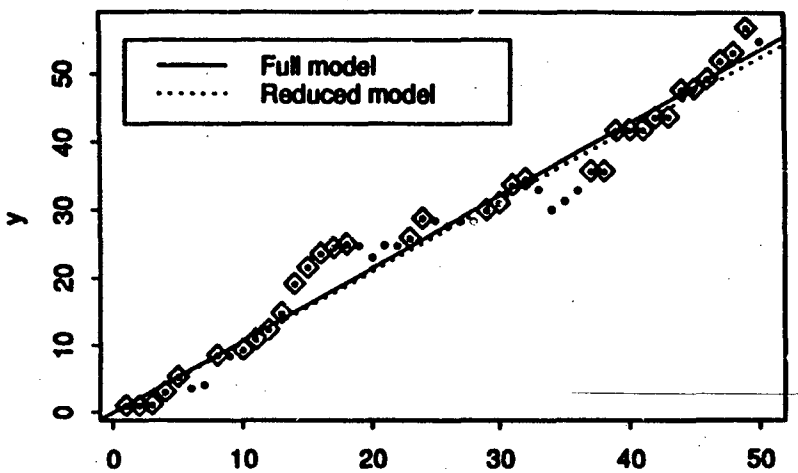
TRYPOS, P. & BLACKMORE, R. (1985). Forecasting records. *J. Am. Statist. Assoc.* **80**, 46–50.

WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics.* Chichester: John Wiley and Sons.

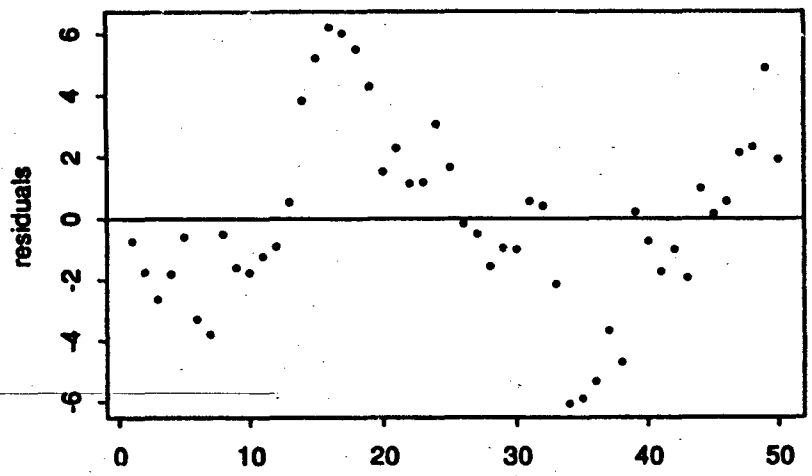
World Almanac and Book of Facts (1989). New York: Newspaper Enterprise Association.

Figure 1. Simulated high correlation record breaking dataset

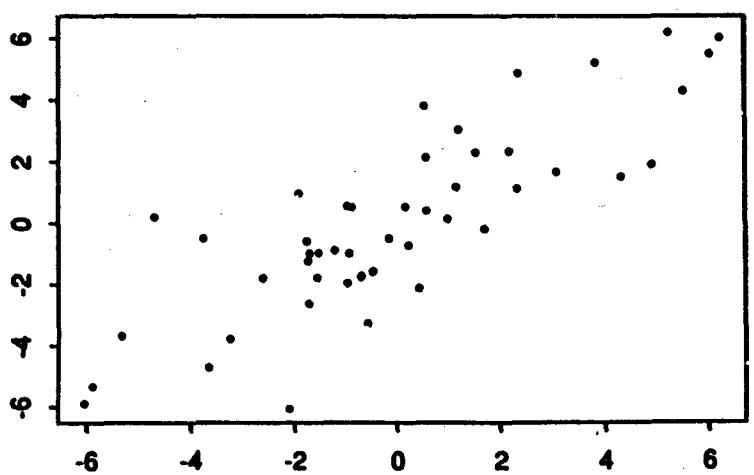
alpha = 0 ; beta = 1 ; sigma = 1.4 ; rho = 0.8



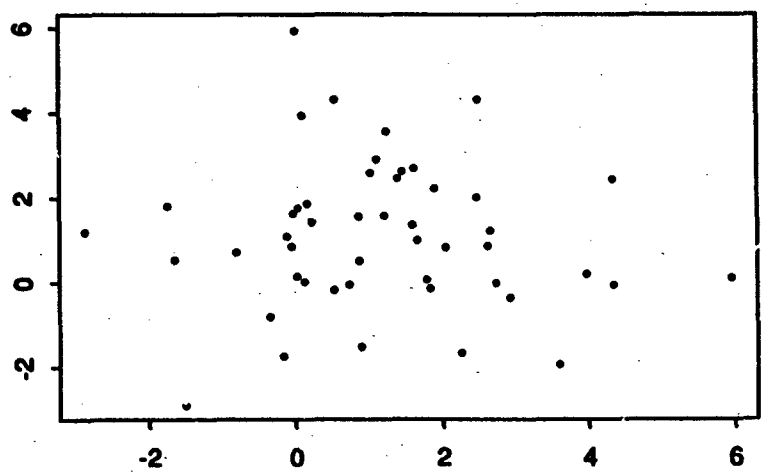
a) Full data with records marked; r= 34



b) Residual plot

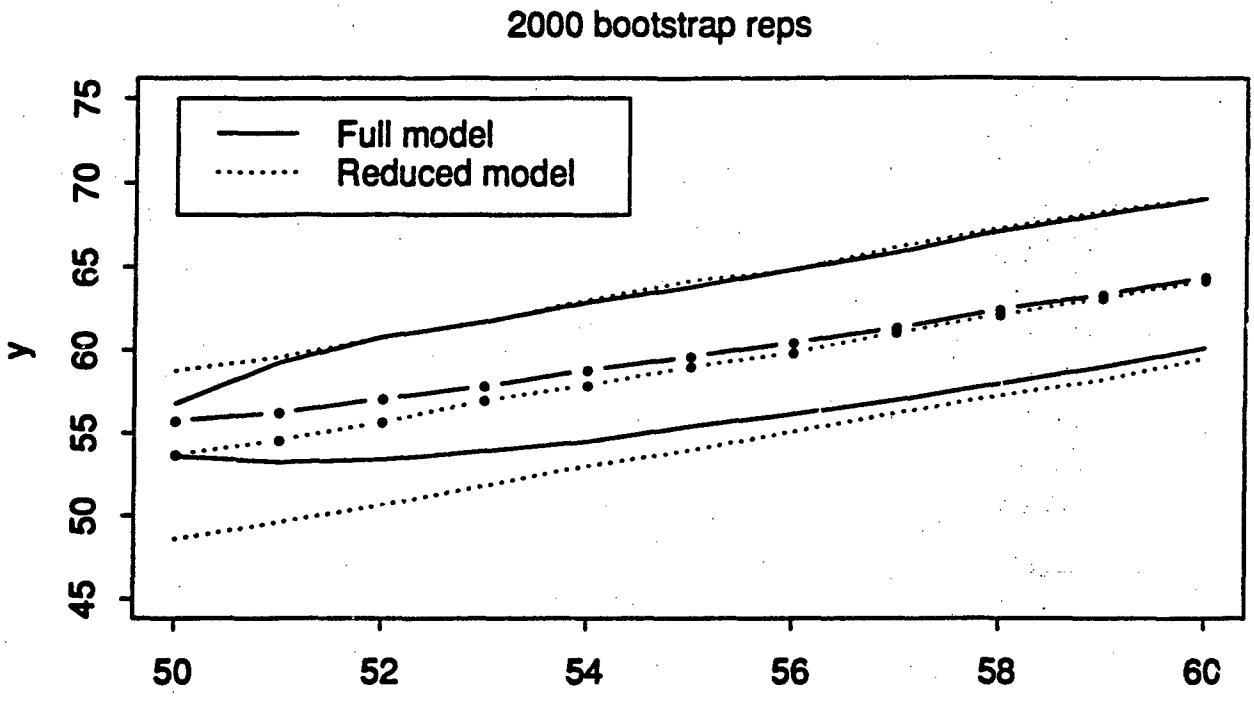


c) Lagged residual plot; corr = 0.83

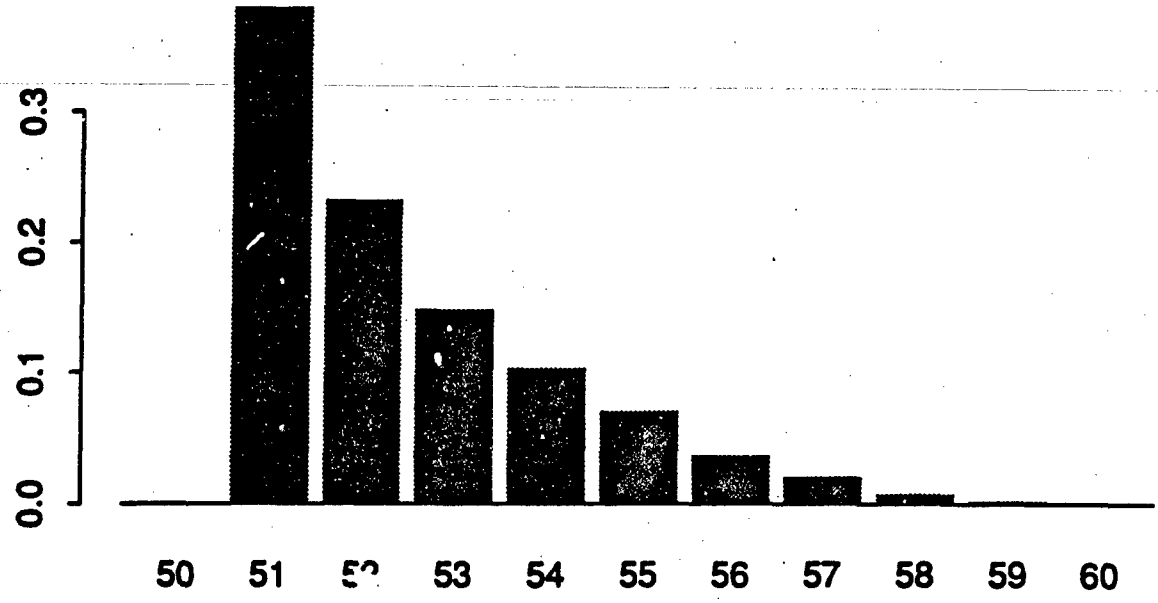


d) Lagged difference plot; corr = -0.05 ; rho hat = 0.901

Figure 2. Prediction for Simulated High Correlation Data

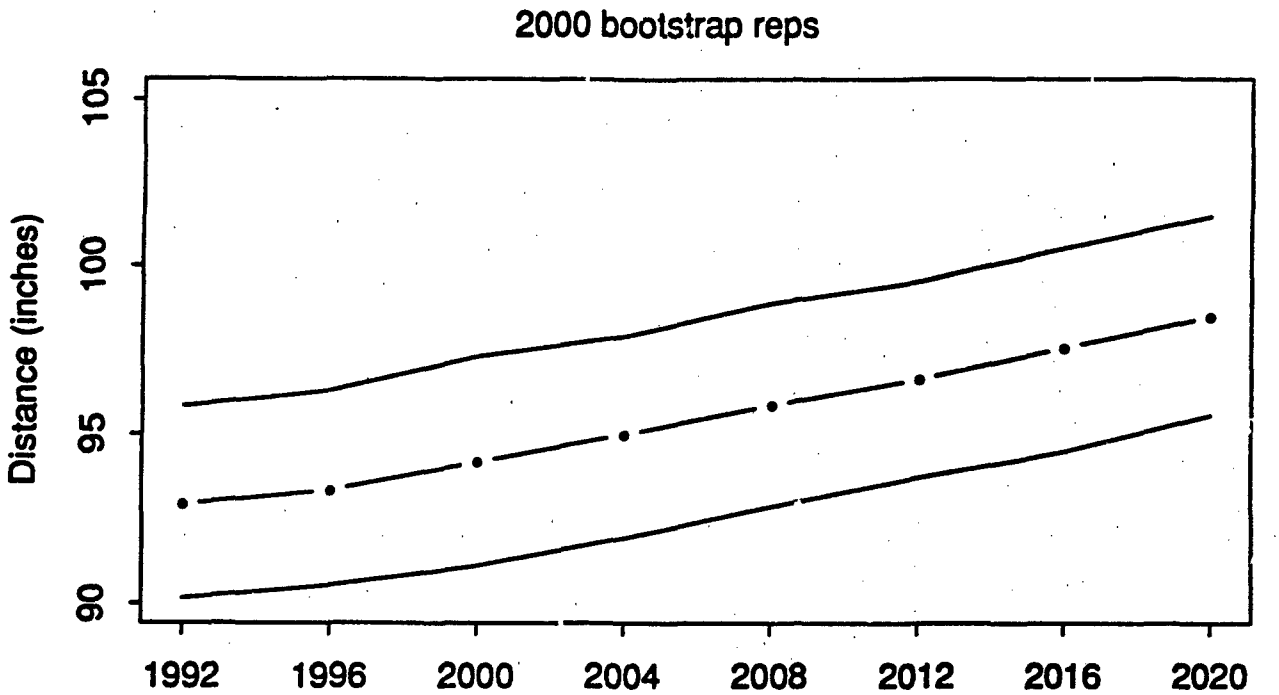


a) 5th, 50th, and 95th percentiles, boostrap distribution of future observations

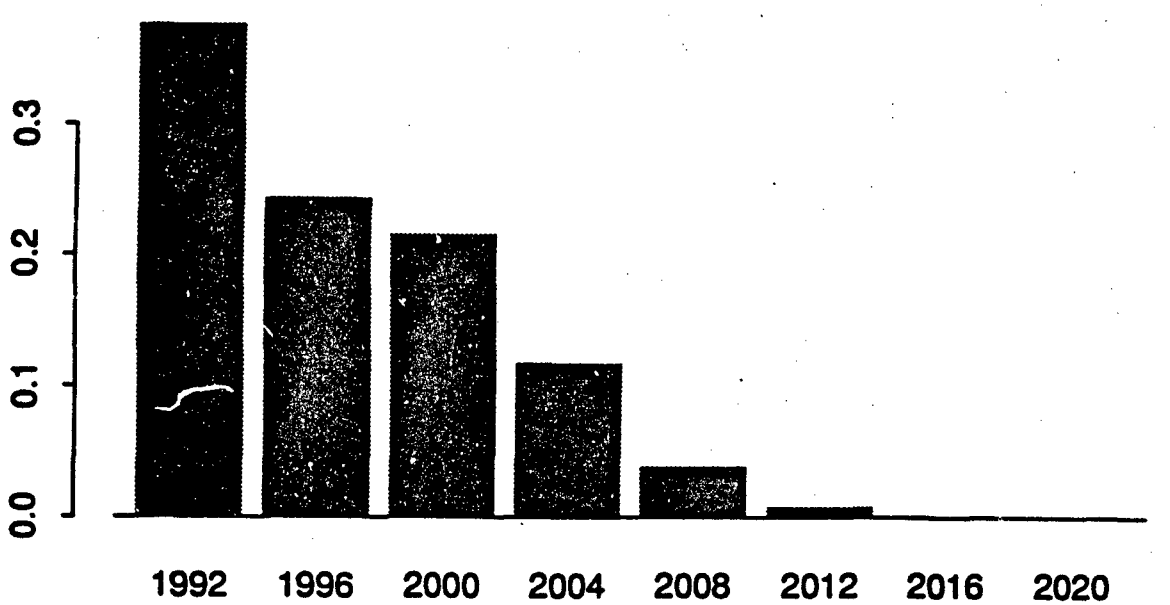


b) Bootstrap distribution of waiting time until next record

Figure 3. Prediction for Olympic High Jump Data



a) 5th, 50th, and 95th percentiles, boostrap distribution of future observations



b) Bootstrap distribution of waiting time until next record

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 465	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Parametric Likelihood Inference for Record Breaking Problems		5. TYPE OF REPORT & PERIOD COVERED: Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): B. P. Carlin and A.E. Gelfand		8. CONTRACT OR GRANT NUMBER(s): N0025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics Stanford University Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research Statistics & Probability Program Code 111		12. REPORT DATE: March 9, 1993 13. NUMBER OF PAGES: 27
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Some key words: Gibbs sampler; Missing data; Monte Carlo approximant.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

Summary

In this paper we consider the analysis of record breaking datasets, where only observations that exceed (or only those that fall below) the current extreme value are recorded. Examples of application areas leading to data of this type include industrial stress testing, meteorological analysis, sporting and athletic events, and oil and mining surveys. The inherent missing data structure present in such problems leads to likelihood functions that contain possibly high-dimensional integrals, thus rendering traditional maximum likelihood methods difficult or infeasible. Fortunately, we may obtain arbitrarily accurate approximations to the likelihood function by iteratively applying Monte Carlo integration methods (Geyer and Thompson, 1992). Subiteration using the Gibbs sampler may help to evaluate any multivariate integrals encountered during this process. This approach enables a far more sophisticated set of parametric models than have been applied previously in record breaking contexts. In particular, we illustrate the methodology for a wide array of discrete and continuous distributional settings, and for observations that may be correlated and subject to mean shifts over time. Related issues in model selection and prediction are also addressed. Finally, we present two numerical examples. The first uses a generated dataset exhibiting a high degree of autocorrelation, while the second involves records in Olympic high jump competition.